Attachment 1

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 31-10-2003	**2. REPORT TYPE** Barchi Slosin	**3. DATES COVERED** (From – To)

4. TITLE AND SUBTITLE A Web-Centric Reference Acquisition and Decision Support System Employing Decisions Theory to Express Relative Preference

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S) Jerald L. Feinstein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) N/A

8. PERFORMING ORGANIZATOIN REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

20031121 085

15. SUBJECT TERMS

16. SECURITY CLASSIFCATION:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
b. REPORT	b. ABSTRACT	c. THIS PAGE		1-20	Jerald Feinstein
Unclas	Unclas	Unclas			19b. TELEPHONE NUMBER (include area code) 202-997-1323

Standard Form 298 (Rev. 8-98) 298-102

A Web-Centric Preference Acquisition and Decision Support System Employing Decision Times to Express Relative Preferences

Professor Jerald L. Feinstein, Ph.D.
The George Washington University
Monroe Hall, 21st and G Streets, NW
Washington, DC 20052
202-997-1323
jerry@gwu.edu or Jerald@post.harvard.edu

71st MORS Symposium
Working Group 25 – Test and Evaluation
10-12 June 2003

ABSTRACT

This paper presents the general theory and techniques involving a web-centric preference acquisition and decision support system employing decision times to express relative preferences or degrees of confidence with respect to decision alternatives. This is an alternative, neural net-motivated method employing decision times or reaction time metrics and a set of decision analytic techniques for capturing, synthesizing, and analyzing decisions, opinions, confidence, and preferences from subject matter experts as well as from the general population.

The paper's first part describes a methodology for transforming the time it takes to make a decision among a number of choices, two-at-a-time, into a set of ratio scaled relative preference or confidence and decision consistency scores using a modified Analytical Hierarchy Process (AHP) approach.

The second part describes a method for synthesizing individual results into a group decision metric, and for assessing the stability of a decision, where stability is the propensity of an individual or group to "change its mind" and defect to alternative choices.

In the third part the focus is on the generality of the approach in terms of how this alternative methodology can be used in different application areas.

In the concluding part, a live, web-centric decision support environment enabling geographically distant decision makers to collaborate in a distributed environment, employing the methodology, is illustrated by discussing screen shots of the application.

INTRODUCTION

The inability of a multiplicity of teams within large organizations to make the right critical decisions, at the right times, and in the right places, is now receiving renewed interest and much needed attention within the Department of Defense, the federal establishment, and in many private sector organizations as well.

According to Peter Drucker (2001), eighty percent of all new products or

services fail within six months or fall significantly short of forecasted profits. Thus, something appears to be very wrong with market research decision making concerning people's preferences for products. Is decision making in the Department of Defense, the federal establishment, or the private sector any better?

Excuses that the challenges and complexities have grown beyond human scale, that available information is often conflicting and ambiguous, or that classic methods do not reflect how decisions are really made can no longer be tolerated due to the enormous and often tragic costs engendered by making decisions that are not only wrong, but not even close.

Decisions are about making choices among alternatives and combining, objective as well as highly subjective information with individual and group expertise. Thus, major decisions in complex organizations involve integrating vast information resources with many choices among a large number of people.

For any complex decision, it is crucial to know if it's a stable one. That is, what is the potential for the individuals involved to switch, back and forth, among different alternatives? How confident are the individuals in the final decision, or to what degree is the selected alternative preferred to all others? How consistent are the individuals in making their decisions among preferred alternatives? The real question is how to address the above challenges in complex organizations with real world constraints to assist decision makers in crafting better and more understandable decisions.

For example, military exercises often contain a Test and Evaluation component requiring the use of subject matter experts to evaluate and analyze results. It has been my experience that the experts' assessments are often exceedingly subjective, with recommendations that are highly sensitive to change, and are strongly linked to the vagaries of which subject matter experts were selected. Often, the experts, themselves, express severe dissatisfaction with the decision process itself. In this paper, the challenges are discussed as well as the degree to which the alternative response latency methodology addresses the problems.

The alternative methodology is motivated by several of the foundation technologies and approaches employed in the field of artificial intelligence - specifically artificial neural networks - and describes a set of decision analytic techniques as well as a web-centric medium for deploying the application among distributed decision makers. The approach addresses some of the challenges for capturing, synthesizing, and analyzing decisions, opinions, confidence, and preferences from subject matter experts as well as the general population. The technique is believed to offer features and advantages not found in other approaches.

The first part of this paper describes the methodology, inspired by biological neuron functionality, for transforming the response latency or the time it takes to make a decision (sometimes referred to as Reaction Time metrics) among a number of choices, two-at-a-time, into a

set of ratio scaled relative preference intensities or confidence scores using a modified Analytical Hierarchy Process (AHP) algorithm. The discussion covers problems with "standard" approaches, known threats to validity, the history of this line of research research, the concept of sensory sampling for highly qualitative information, and research comparing this approach to a classic method in terms of a gold standard.

The neural connections to decision making have been studied and reviewed extensively [Glimcher PW, 2001], [Gold JI, Shadlen MN, 2001], [Platt ML, 2002], [Romo R, Salinas E, 2001], and [Schall JD, 2001]. Many fascinating models and conjectures have been proposed, so the interested reader can explore recent work within the selected references. For an excellent background on an exploration of the connections between reaction time and neuron-based processes, the reader is directed to a very recent paper by Schall (2003).

While the general concept of neuron-based systems and sensory sampling provided the early motivation to explore the notion that complex decisions employ neural processes, and the more complex the decision, the more processing and time are required, it is important to understand that the references cited are not required to be understood before the reader can gain a substantial understanding of the techniques discussed later in the paper.

In the second par of the paper discusses methodologies for synthesizing individual decision preferences into a group metric, and for clustering results across the alternatives to develop a dynamic consensus map for assessing the stability of a decision. I define decision stability as the propensity of an individual or group to "change its mind" and defect to, or oscillate among, alternatives. State variables and linked differential difference equations are employed to simulate movement of individuals among the decision states employing relative decision preference intensities as transition propensities to assess stability and make forecasts about the likelihood and directions toward which decisions might change.

The third part focuses on the generality of the approach in terms of how this alternative methodology generates metrics and information useful in areas different application areas such as linear and integer programming, game theory, cost analysis, capital budgeting, proposal and program evaluation, modeling and simulation, evaluation of military exercises, and a host of other applications.

In the concluding part, a live, web-centric decision support environment enabling geographically distant decision makers to work in a distributed environment employing the methodology is illustrated by discussing screen shots of the application.

The core application resides on a central server in British Columbia linked to a database from which decision alternatives are presented, two at a time, to decision makers at any location with web access. Alternatives are selected by clicking within the appropriate area within a browser window, and the choices selected as well as decision times are returned to the server and stored appropriately within the remote database. Finally the server side software uses modified Analytical

Hierarchy Process (AHP) scaling algorithms to generate the ratio scales or priorities characterizing the decision maker's relative preferences or confidence about each alternative as well as a measure of consistency over the alternatives.

GENERAL

Currently, an approach employed to make assessments about the effectiveness of military exercises with Test and Evaluation components is to ask experts to self-report their degree of preference or confidence that they associate with their recommended assessments [Harmon and King, 1985]. One common method of self report is to assess scores based on relative measures of performance of different military systems using a 1-9 scale or similar method. Employing the self report approach, experts must report not only their preferred recommendation; they also must provide an estimate of the degree of preference or confidence for their recommendation. Because these results are generated from experts' self-reported verbal or written assessments, there is some concern about criterion validity.

Problems and Threats to Validity

Reporting the preferred recommendation is a relatively easy task; however, quantifying the degree of preference or confidence in one recommendation over others is something that almost anyone finds to be a terribly difficult and distasteful undertaking. This is because they are being asked to quantify something that they do not normally quantify in practice. While providing a recommendation based on a set of conditions can be accomplished relatively quickly; quantifying relative levels of preference or confidence among alternatives takes much longer, and the results are often suspect for reasons discussed later in this paper.

To further complicate the situation, conscious processes are involved when deciding on relative levels of preference or confidence, and these processes are known to be vulnerable to manipulation. People often respond to the need for a decision in a manner that they think pleases someone, or respond in ways that tend to maintain a positive image.

In my interviews with subject matter experts and decision makers working within large organizations, I discovered that several very real threats to effective decision making were obvious to these people. They knew very well why bad decisions happened in their organizations. They claimed it had to do with the organizational context in which they worked. Reasons provided were that making certain good decisions was often viewed as a bad career move. Two other related reasons were individual and group cognitive dissonance, another two were psychotic management or organizational neuroses, and finally groupthink which is related to the idea of group cognitive dissonance.

Thus, the self-report approach is not only time-consuming; it is known to be vulnerable to conscious censure and manipulation [Clemen, 1996].

To make matters worse, decision makers' self-reported preferences or estimates of confidence can be inconsistent and may or may not be a good measure of their true feelings. For example, decision makers may be consciously unaware of their true feelings [Banaji and Greenwald, 1994; ;Nisbett and Wilson, 1977], or they may

be reluctant to reveal their true feelings [Crosby, Bromley, and Saxe, 1980; Gaertner and Dovidio, 1986; Sigall and Page, 1971].

Self-reported information may be subject to further degradation as subject matter experts can become annoyed with the time consuming and often uncomfortable process of not only responding to decision makers' questions about which is the preferred alternative, but being asked, repeatedly, to estimate the degree to which the final decision is preferred to other alternatives [Marshall and Oliver, 1995]. Subject matter experts often develop anxiety over the requirement of directly reporting a degree of preference [Medsker, 1998], and if they are uncomfortable with the feeling that they might provide an incorrect answer, their certainty may waver [Medsker, 1998], or they may view the process as a waste of time [Marshall and Oliver, 1995] and provide inaccurate responses.

There is some unease about peoples' self reported opinions, even when they are very confident of their recommendations. This is because it is reported that people, in general, and some in particular, have been found to use heuristics or "logic" that are known to be error-prone and produce systematic errors in both their recommendations as well as estimates of confidence associated with the recommendations [Kahneman, Slovic, and Tversky, 1999].

Lam [1998] states that knowledge and feelings of confidence about recommendations that can be expressed in words and symbols represent only the tip of the iceberg of the entire body of possible knowledge. The idea that some components of knowledge are tacit, sometimes referred to as subjective or unconscious, presents a problem that a decision maker's degree of confidence in, or degree of preference for a tacit recommendation is difficult, if not impossible, to express verbally. This idea was first mentioned by Polanyi [1962 and recently by Gerald Zaltmen (2003), a Fellow at Harvard University's interdisciplinary Mind, Brain, Behavior Institute. He states that 95 percent of thinking happens in our unconscious. Tacit knowledge refers to knowledge that is intuitive, unarticulated and that cannot be easily codified and transferred. Polanyi stated that "We know more than we can tell", and maintained that a large part of human knowledge is occupied by knowledge that can never be articulated.

Thus, such concerns about the criterion validity of verbal or self-reported codified information underscore the problem and the overwhelming need for developing more accurate, less time consuming, and more unobtrusive measures [Dovidio and Fazio, 1992] which will help to mine knowledge with tacit components more effectively than is done currently.

A Possible Solution Employing Response Latency

The time it takes to make a decision, sometimes called response latency or reaction time, may provide a solution to the self reporting problems because it is known that the faster a choice is made between two alternatives, the more the selected alternative is preferred to the other and the stronger the relative certainty of the decision. This notion appears to be well founded.

A substantial body of literature supports response latency as a measure of

preference and certainty in psychological research and market analysis studies [LaBarbera & MacLachlan, 1979]. However, its methodological use in a structured decision making paradigm appears almost nonexistent.

Early work by Dashiell [1937], Cartwright [1941], Festinger [1943], and Tyebjee [1979] provides evidence that response latency tends to be inversely related to the degree of certainty experienced in making a judgment and inversely related to the objective or subjective distances between the choice-pairs; that response latency could be used to estimate the difference in affective value across multiple decision variables, and it is insensitive to the presentation order of the choice-pairs.

Sensory and Subjective Sampling

Try this experiment if you like, or just think it through. If we have two balls, or other small objects that appear identical other than they have different weights, and place one in each hand, as illustrated in Figure 1. Subject testing the weights of different objects.

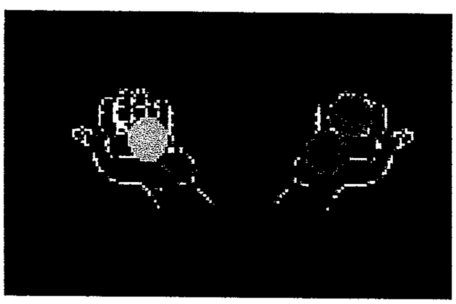

Figure1. Subject testing the weights of different objects.

Now you are asked to judge which one is heavier. If one is much heavier than the other, that means the effect size is large, and we tend to respond very quickly with the appropriate answer. On the other hand, if one object is only slightly heavier than the other, it takes longer to decide. When I conduct this experiment with audience volunteers during presentations, the result is that, for small differences in weight, the subject, who is assessing the weights, always seems to "jiggle" the objects. When asked what they are doing when they are "jiggling" the objects, they often respond with phrases such as "testing the difference in weights."

Basically, when we are performing the "jiggling" action, we are engaged in sensory sampling. When there is a very small difference in weights (small effect size), it takes longer to decide since more samples are needed in order to differentiate between the two. If there is a large difference in weights, fewer samples are required for a given level of confidence, and we can decide very quickly. This is conceptually similar, to the classic hypothesis test from statistics, however, the underlying mathematics is different. Thus, the decision time is an inverse function of the difference in relative weights.

In another example, familiar to anyone who has ever experienced an eye examination to determine the correct lens prescription knows the procedure begins by the optometrist asking the person being examined to look through a device at a set of images illustrated in Figure 2. An eye examination.

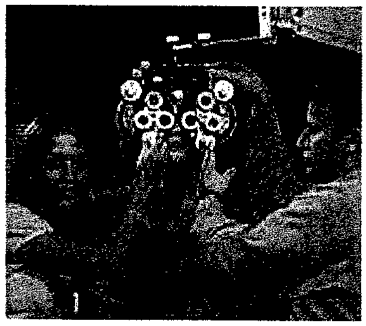

Figure 2, An eye examination

The images are presented two-at-a-time, and then the specialist rotates the lens settings and asks which alternative is clearer? In the beginning, when there is a substantial difference in clarity between the two choices, one can answer very quickly. Then, as the exam progresses, the differences between the clarity of the two images becomes less and less, and it takes much longer to judge which one is clearer. Often when the differences become very small the patient must ask the optometrist to flip back and forth so as to gain even more information by taking additional samples on which to base the judgment. In summary, the less the difference, the more sensory samples are required.

Since variables such as weight, the intensity of light, and the clarity of an image can be linked to a realm of reality that some refer to as objective and can be sampled by the senses, we say that sensory sampling has a strong link to objective reality.

However when we deal with affective variables which are those related to attitudes and feelings, and include interest and motivation, how we feel about something, likes and dislikes, "gut" feelings, and tacit knowledge; we understand that the variables represent measures of high subjectivity, yet often of a substantially higher value than the former. In this case, the decision maker can be thought of as subjectively sampling internal information archives. This concept is illustrated in Figure 3. How subjective sampling works.

Subjective Sampling Taps Deeper "Gut" Feelings and Emotions About Decisions

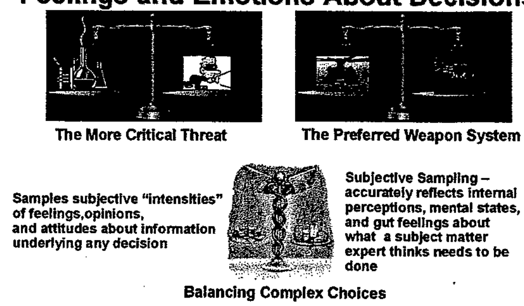

Figure 3. How subjective sampling works.

For example, when we choose between two very similar subjective alternatives, using a metric such as an assessment of beauty, the decision time will be longer than when we select among alternatives that are far apart on the beauty scale.

Thus, this paper reports on the results of research designed to describe and compare an alternative approach which is possibly more efficient at knowledge elicitation and decision support, and is based only on a decision maker's selected recommendation and response latency or decision time. The goal is a better method for assessing a decision maker's or subject matter expert's feelings of preference, confidence, or certainty in their judgments and recommendations.

In this alternative approach, the inverse of response latency is used to estimate an expert's degree of relative certainty or confidence in the selected choice in a paired comparison of possible recommendations. The response latency method has the

advantage of being unobtrusive, less prone to conscious censure, quicker to perform, requires less effort, and is less expensive to administer.

A Key Idea - Combining Response Latency and Analytic Hierarchy Process (AHP) Scaling Methodology

Saaty (1977) introduced a method of computing relative weights of a positive pairwise comparison matrix of judgments by the eigenvector method, and is summarized from the references as follows: let **A** be the positive pairwise comparison matrix with respect to *n* criteria that is illustrated as follows:

$$\mathbf{A} = \begin{bmatrix} w1/w1 & w1/w2 & \cdots & w1/wn \\ w2/w1 & w2/w2 & \cdots & w2/wn \\ \vdots & \vdots & \vdots & \vdots \\ wn/w1 & wn/w2 & \cdots & wn/wn \end{bmatrix}$$

Where Wa/Wp represents ratios of the "a"th factor of importance, degree of preference, or related factor over the "p"th factor. $(a, p \in 1,2,...,n)$ and in the classic approach, the value of Wa/Wp is also a subjective assessment of relative value by the decision maker.

Multiplying **A** by the vector of relative importance or preference vector: $W = (W_1, W_2, \dots W_n)^t$ results in the equation:

$$\mathbf{AW} = \begin{bmatrix} w1/w1 & w1/w2 & \cdots & w1/wn \\ w2/w1 & w2/w2 & \cdots & w2/wn \\ \vdots & \vdots & \vdots & \vdots \\ wn/w1 & wn/w2 & \cdots & wn/wn \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix}$$

The method uses the maximum eigenvalue to find the general **W.** Then, a set of linear equations for $W_1, W_2, \cdots, W_n$ can be obtained, and by adding the normalization constraint, $W_1 + W_2 + ... + W_n = 1$, exact values of $W_1, W_2, \cdots, W_n$ can be calculated.

The process generates a set of relative preference or confidence intensities for all alternatives, and these are called priorities. As discussed, they are derived from the elements of an eigenvector formed from the matrix of a decision maker's judgments based on pairwise comparisons of options or alternatives. Also provided by the AHP approach is a measure of judgmental consistency, where consistency is a quality of the elements of the judgment matrix of pairwise comparisons where the transitive property is maintained among the matrix elements such that $a_{ij} = a_{id} \times a_{de} \times \dots \times a_{mj}$. The measure of consistency is called the Consistency Ratio [CR], which is actually a measure of inconsistency, and is defined by the following formula:

$$\mathbf{CR = \lambda_A - order\ [A] \,/\, [order\ [A] - 1]\ \lambda_{random}}$$

where λ_A is the largest eigenvalue of the reciprocal matrix of pairwise comparisons A, and λ_{random} is the largest eigenvalue of a randomly generated reciprocal matrix of the same order as matrix A, and Order [A] is the order of matrix A. In practice, obtaining a Consistency Ratio of greater than .1 is considered to be a high level of inconsistency and cause for closely examining the structure of the decision.

In summary, this process provides a method for generating a ratio scale of an

individual's judgments in terms of relative levels of certainty or preference over a set of alternatives or options using pairwise comparisons. That is, a decision maker compares, all options, two-at-a-time, and in the classic AHP approach, selects one of the two options and associates it with an integer between one and nine, inclusive. The integers represent the decision maker's judgments in assessing degrees of preference or confidence for the options selected over those not selected in the pairwise comparisons. A "one" equates to about equal preference or confidence and a "nine" indicates the greatest degree of preference in the alternative selected over the one not selected. For a detailed description of this technique, the reader is directed to Saaty, [1994,1995].

However, in our approach, we modified the classic AHP technique where we have replaced the 1-9 range of assessments with an inverse function of the decision maker's decision time or response latency to indicate degrees of preference and confidence in one choice over another. Thus, functions of time form the inputs to the eigenvalue process, and one advantage is that the method normalizes all responses on a common scale such that if one respondent has a slow response and another is typically quick, those differences are addressed within the normalization process.

Summary of the Approach

Now we illustrate the combined response latency AHP methodology by discussing a concrete numerical example.

A dataset from one decision maker is used as an example to illustrate the modified AHP method for using decision times to calculate relative degrees of preference for alternatives that are expressed as ratio scaled priorities in addition to a consistency ratio for the decision set.

For each choice-pair presented the decision maker decides on the preferred alternative by clicking on the desired choice within the browser window using our web-centric decision support environment. After the decision maker responds, both the selected alternative and the time taken to respond are captured in the browser and transmitted to a remote server where it is stored within a central database. Response latency is measured in increments of hundredths of a second. If the latency is greater than nine seconds, a value of nine seconds is used in the analysis. This level of truncation is of the same order of magnitude used in other research [LaBarbera & MacLachlan, 1979].

Next, the inverse of the response latency to select an alternative is calculated and rescaled using dimensions of 1/dekaseconds, and this transformation is used as the appropriately scaled raw measure of relative preference on a range similar to the classic approach. For all choice-pairs, this information is stored within a matrix of pairwise comparisons. Then, as prescribed in the AHP, if the column variable dominates the row variable within the matrix of pairwise comparisons, the inverse of response latency is calculated and inserted into that matrix element; if the row variable dominates the column variable, then the response latency [in dekaseconds] is calculated and inserted into that matrix element. Dominance is determined by the choice selected from

the choice-pair presented. The procedure is illustrated by employing actual data collected from a decision maker as shown in Figure 4. Matrix of Pairwise Comparisons, Subject 5, for the Response Latency Method.

The rows and columns are the investment choices offered: CIH – Cash In Hand, AGF – Asian Growth Fund, SPI – S&P Index Fund, BCF – Blue Chip Fund, and MMF – U.S. Government Money Market Fund. Each matrix element, a_{ij}, represents the decision maker's preference in selecting choice i over j. If "i" is preferred to "j" then the rescaled value is used (1/dekaseconds), else the response latency in dekaseconds is used.

As prescribed in the AHP, the elements below the diagonal are the inverses of corresponding values above the diagonal [$a_{ji}=1/a_{ij}$]. In example illustrated, AGF is chosen over CIH, and a_{12}=.206, which is the value for the response latency in dekaseconds. That signifies that the subject was about 1/5 as certain about CIH than about AGF. However, SPI is chosen over MMF and a_{35}=4.27, which is the inverse of the response latency in dekaseconds. This means that the subject was more than four times more certain about SPI than about MMF.

The process works like this, first, CIH is compared with each variable along the row - first itself, then, AGF, next, SPI to MMF. If CIH is selected over the variable it is being compared to, then the inverse is used; if they are the same, a one is entered (as along the diagonals), and if the other variable is selected, the response latency in dekaseconds is used. Next, AGF is compared to itself, then SPI, to MMF. Each element above the diagonal is entered in this manner, and the elements below the diagonal are entered as reciprocals of those already entered above the diagonal as is indicated above. As mentioned earlier, the resulting matrix is illustrated as Figure 4. Matrix of Pairwise Comparisons, Subject 5, for the Response Latency Method.

	CIH	AGF	SPI	BCF	MMF
CIH	1	.206	.241	.354	.565
AGF	4.854	1	.371	.211	.276
SPI	4.149	2.695	1	.428	4.274
BCF	2.825	4.739	2.336	1	4.367
MMF	1.77	3.623	.234	.229	1

Figure 4. Matrix of Pairwise Comparisons, Subject 5, for the Response Latency Method

Next an eigenvalue method is used to calculate a set of priorities from the response latency data by employing the iteration approximation of raising the matrix to higher and higher powers. In addition to the priorities, a Consistency Ratio is calculated also.

After using the eigenvalue technique to calculate Subject 5's assessment of confidence [shown as priorities] for the different alternatives as well as the associated Consistency Ratio are illustrated for a BOOM economic scenario as illustrated in Figure 5. comparing a single subject's priorities as Certainty Factors and Consistency Ratio [CR] for response latency and self-report methods.

Note that Figure 5 contains assessments of confidence and consistency for both the response latency and self-report method for the same subject. Also note that the priorities associated with the subject matter expert's relative

preferences for each choice are different depending on which approach is used as are the consistency ratios.

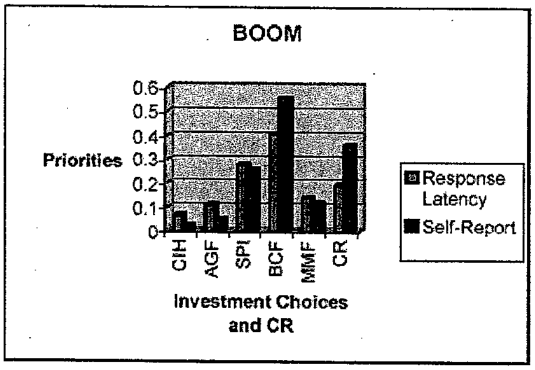

Figure 5. Comparing a single subject's priorities as Certainty Factors and Consistency Ratio [CR] for response latency and self-report methods.

This is an important observation because it is crucial to determine which set of calculated results is better using a gold standard. Since we captured both the decision times and the self-report 1-9 scale assessments from the subjects at the same time we constructed two hypothesis tests in order to assess which approach was better. The first test needed to determine if the evidence was sufficient to reject the null hypothesis that the two approaches produced the same sets of priorities. If not, then a second hypothesis test was used to determine if there was sufficient evidence to reject the null hypothesis that the mean consistency ratio for the self-report was higher than that for the response latency method. Note that low consistency ratios mean higher consistency among the assessments; and therefore, they are more desirable than higher consistency ratios.

Our assumption is that since consistency is defined as the degree to which a highly structured relationship exits among the individual elements of the matrix of pairwise comparisons such that the transitive property is maintained, it is unlikely that this ordered relationship among the elements of the matrix would arise purely by chance.

The logic is based on the idea that if there is some underlying consistency in the decision maker's assessments of confidence on different alternatives, then the preferred methodology would be the one which detects those consistencies better than the other. A rationale for this approach is that if we view the same process through two different instruments, the instrument defined as better is the one that detects consistency better than the other.

Summary of Results

To assess statistical significance, paired sample t-tests were employed, and sufficient evidence was found to reject the null hypotheses ($\alpha_. =.025$), N=41, that the mean of the difference in the confidence estimates between the two methods was zero ($p < .0001$). This means that we can be reasonably confident that the response latency and classic self-report methods produce different sets of confidence estimates.

Next, we need to assess which method captures the underlying consistency best. In a second test, it was found that there was sufficient evidence to be at least 95% confident of rejecting the null hypothesis and accepting an alternative hypothesis that response latency detects a lower consistency ratio than does the self-report method ($p < .0001$).

Thus, it was found that the two methods produce different sets of confidence estimates and that the response latency method produces a lower Consistency Ratio (detects consistency better) than does the self-report method. That is, when consistency is present, the response latency approach appears to detect it better than the self-report approach, and that is our gold standard.

In assessing practical significance it was found that a substantial percentage of response latency Consistency Ratios are lower than the ".1" threshold identified in the AHP theory as meaningful. That is, Consistency Ratios larger than .1 are cause for further examination, and possible concern. On the other hand, a very small percentage of self-report Consistency Ratios were found to be beneath the threshold. This means that a large proportion of confidence estimates derived from the response latency approach meet the .1 test used in the AHP; however, the opposite is true for estimates derived from the self-report approach. This is a question of practical significance. These findings are illustrated in Figure 6. Comparing classic and response latency approaches for practical significance.

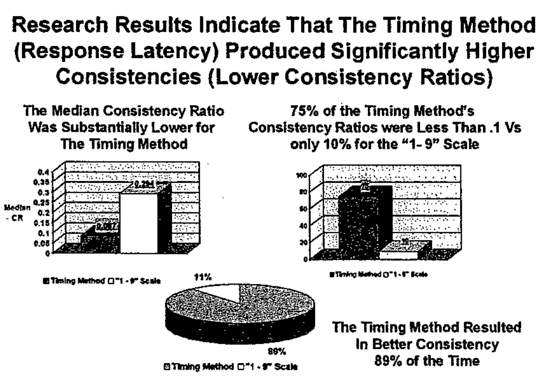

Figure 6. Comparing classic and response latency approaches for practical significance.

In examining the histogram of the Consistency Ratio for response latency, one finds a median of .087, which is less than the .1 threshold, with values of less than .1 considered to be a good indication. More importantly, approximately 75% of Consistency Ratios derived from response latency fall between .119 and 0.

However, the Consistency Ratio for self-report samples is distributed differently. The histogram for the these has a median of .294, significantly greater than the .1 threshold value, with values greater than .1 considered to be a poor indicator and reason for possible rejection. Further, only 10% of these samples fall between .123 and 0. Further, for the self-report method, approximately 90% of the cases fall above the .1 threshold which is cause to examine closely their validity, since for the response latency approach only 25% of the Consistency Ratios fall above the same threshold.

Thus, the differences found are not only statistically significant; there is some tentative evidence that the results are highly practical. Additional testing on a broader range of subjects and in different knowledge domains needs to be conducted in order to make any substantial claim in this area.

Some Interesting Findings

Possible conscious intervention was detected in many subjects, and the data collected for Subject 9 is used to illustrate the point. It was discovered that the subject entered a "nine" as an expression of ultimate confidence in the decision for every case of choice-pairs. The result was a high Consistency Ratio, near .5, (a high consistency ratio means low consistency, and in this case a .5 is almost off the scale, since a level of

greater than .1 is cause for some concern). At first, one might suspect that the subject was exhibiting a set response. However, when the response latency results are examined one discovers a rich variation in the response times and a set of confidence estimates with a Consistency Ratio below the .1 threshold. Evidently, the subject was not responding casually to the question of which choice is preferred. However, when the subject responded to questions concerning the degree to which one choice is preferred to the other, it was found that all responses were the same and reported a very high degree of confidence for each response. Thus, the self-report method yielded suspicious results with a high level of inconsistency while the response latency approach resulted in more plausible results with a much lower level of inconsistency present for the same individual and same decision set..

This result appears to be a classic example of what the literature predicts. For instance, did the subject exhibit conscious censure where a degree of confidence less than "almost certainty" was not acceptable? Was there some annoyance or uneasiness at the difficulty of trying to quantify degrees of preference that the subject was not accustomed to quantifying? Did the subject try to quickly end an uncomfortable process? All of these factors represent noise that can interfere with the accurate detection of assessments and result in lower levels of consistency. Yet, it is in just these types of cases that the response latency technique is proposed as a remedy. In this case the difference is very large in that the self-report results in a high level of inconsistency with a Consistency Ratio near .5, while the response latency approach results in a low inconsistency with a Consistency Ratio of less than .1

Summary of Section One

Based on this research, one can conclude that there is evidence to support the use of response latency, combined with the AHP, as a method for eliciting and scaling assessments of relative preferences and confidence from decision makers, subject matter experts and others. It appears that the response latency method offers a more accurate method that is easier to use, elicits information more rapidly, is less costly, and requires less effort in that the information is obtained unobtrusively and with higher confidence. Concerns about the criterion validity of self-reported information underscore the problem and the need for using more unobtrusive measures [Dovidio and Fazio, 1992]. Since this approach collects response latency in an unobtrusive manner, no additional effort is expended by the decision maker or subject matter expert in articulating levels of preference or certainty and has no knowledge that decision time information is being collected.

Since relative preference or confidence data are collected without any additional effort on the part of the subject, the elicitation process should proceed much quicker. For instance, in the response latency case it takes less than ten seconds to state a preference, where in the self-report approach the subject still must state a preference, and then try to quantify a degree of preference. The additional effort tends to increase the time used by more than one order of magnitude.

The approach selected to evaluate the two methods was based on the possibility that each method would generate different sets of Consistency Estimates, and the task was to assess the degree to which each method detects each subject's consistency across the choice-pairs.

Applications

The notion of being able to translate highly qualitative "gut" feelings and tacit knowledge into ratio scale metrics has wide application in a number of fields currently being explored. Several applications will be discussed briefly to provide the reader with a broad understanding of how this technique might be used in practice.

One application area of interest is computer security threat analysis and planning. Here the potential threats are identified; however what is not known are the relative danger, system vulnerabilities costs to correct, and a host of other related variables.

Be employing subject matter experts to assess relative levels of threat, likelihoods, and vulnerabilities so that we may develop a rigorous plan to minimize the threat for the lowest possible cost or to learn what can be accomplished for various levels of expenditures. After generating the relative threat, likelihoods, and vulnerabilities, other subject matter experts with the cost analysis domain are employed to assess relative levels of costs for each mitigation initiative.

Next, the relative improvements are divided by the relative costs to implement each improvement. This provides a method to employ the capital budgeting, mixed integer programming to generate optimum initiative mixes to correct deficiencies. This general approach has wide application in other areas including weapons systems analysis, general cost analysis, acquisition studies, and related initiatives. The critical part is that absolute measures are not required, only relative assessments that are used to generate the ratios.

The technique also has wide application in the political forecasting and intelligence analysis areas; particularly HUMINT. We have had substantial interest from the large political consulting firms, advertising agencies, management consulting firms, and other sources for these kinds of applications. Consider that the technique may be used as a survey tool to quickly access the opinions of large numbers of people, and that information can be rapidly brought together in consensus maps for illustration and stability forecasts. An interesting feature of the approach is the consistency ratio measure for each person making an assessment. In the general population we hay find numbers of people who do not care, get distracted, do not know anything about the domain, or understand neither the questions nor context. In these cases, the consistency ratio for a respondent will be high, and it is possible to purge these individuals, based on some threshold level, from the calculations.

The response latency approach has wide applications within any test and evaluation exercise where subject matter experts are used to evaluate and assess results. In these assessments small numbers of subject matter experts are used often as a matter of convenience.

In these instants it is crucial to employ consensus maps to identify situations where the assessments possess low stability. It is in those kinds of situations, under many contexts, that actions that were based on the assessments were tragically wrong.

In any group decision context where it is important to drive consensus, the response latency approach offers interesting features such as tracking stability and group consistency.

In other disciplines such as game theory it is crucial to understand subjective values that are placed on gains. In more interesting areas such as in multiplayer scenarios, each can have different subjective assessments of the environment.

In acquisition management and proposal evaluation, this approach offers easier assessments of alternatives. In one application we worked with a firm who used the method for employee performance evaluations and assignment of bonuses.

These application areas provide only a brief introduction to the different uses anticipated by those having an understanding of the technique.

Synthesizing Results and Stability Assessments.
To illustrate the method for synthesizing results from individual subject matter experts into a group decision metric we propose an illustrative example where a group of intelligence experts assess the threats from different terrorist organizations. For this application the organizations will reflect the names of their leader and is a purely fictional example. The organizations considered are those run by Attila, Azov, Po, Qudzu, and Karif.

In this simple example, we ask the intelligence experts to assess the relative threat levels of the different organizations by employing the web-centric decision support environment as described in Figure 7. Steps used to demonstrate threat assessment.

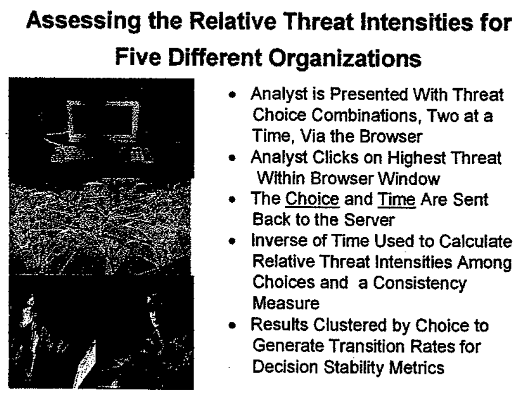

Figure 7. Steps used to demonstrate threat assessment.

To simplify our discussion, we will limit our initial illustration to three of the organizations and later show data representing all five.

The three organizations selected for the initial discussion are Attila, Po, and Azov. To develop the dynamic consensus map, the individual analysts results are clustered by choice so all analysts who selected Azov as the most serious threat are put in one cluster and median threat intensities and a Consistency Ratio are generated for that group. The same is done for each of the three groups and the results are graphically illustrated in Figure 8. Dynamic consensus map.

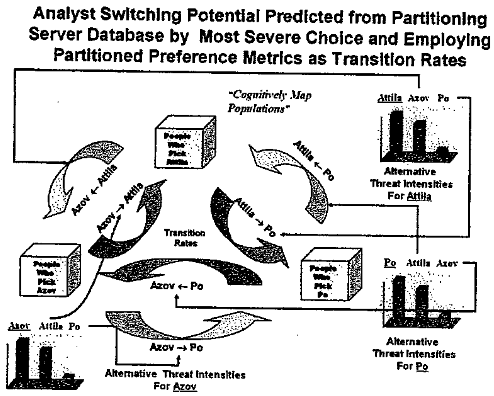

Figure 8. Dynamic consensus map

Each box represents a preference state containing the number of analysts who selected that organization as the most severe threat. For those in each box, we use their individual threat estimates or priorities that were generated using the eigenvalue method and synthesize them into a group median. So that we now have a set of primary and alternative threat intensities associated with each box or state represented by the bar charts adjacent to each box. The intensities are all normalized si that sum to one, and represent the propensity for an analyst to switch from the current opinion to another and the conjecture is that they function similar to transition rates in Markovian processes.

A state variable modeling approach employing analyst population counts and "strength of preference" metrics to generate stability forecasts is now discussed. A canonical model for the Attila, Azov, and Po example is illustrated next. The K_{ij}s are derived from "strength of preference" metrics, and the N_ks are the numbers of analysts supporting a particular assessment.

For Example: $N(t)_{Atilla}$ represents the number of analysts who selected Attilla over all other choices at some point in time, and $K_{Attila \to Azov}$ represents the potential for those analysts to switch from Attila to Azov and is derived from the preference metrics.

The set of simultaneous differential difference equations that expresses the relationship are given by:

$N_{Attila}(t+1) = N(t)_{Attila} - (K_{Attila \to Azov} + K_{Attila \to Po})N(t)_{Attila} + K_{Azov \to Attila}N(t)_{Azov} + K_{Po \to Attila}N(t)_{Po}$

$N_{Azov}\,(t+1) = N(t)_{Azov} - (K_{Azov \to Attila} + K_{Azov \to Po})N(t)_{Azov} + K_{Attila \to Azov}N(t)_{Attila} + K_{Po \to Azov}N(t)_{Po}$

$N_{Po}\,(t+1) = N(t)_{Po} - (K_{Po \to Azov} + K_{Po \to Attila})N(t)_{Po} + K_{Azov \to Po}N(t)_{Azov} + K_{Attila \to Po}N(t)Attila$

$N_{Atilla} + N_{Azov} + N_{Po} = \text{some } K$

Employing these kinds of relations we are developing constructs to forecast the degree to which decision sets remain stable and describe their ranges of variation.

The transition matrices for all five organizations are illustrated in Figure 9. Relative threat assessment transition matrices.

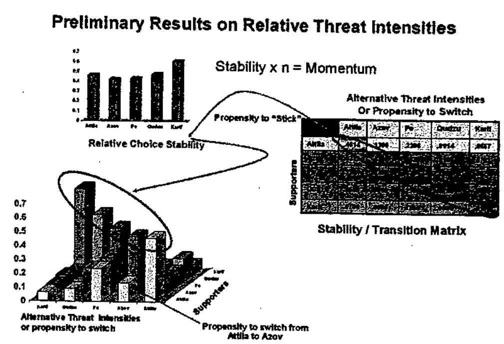

Figure 9. Relative threat assessment transition matrices.

The Web-Centric Decision Support Environment

This section provides a brief view of the web-centric decision support environment in terms of screen shots of a trivial example where a subject matter expert on football is asked to assess team quality. Some of the details of the environment were summarized earlier.

First, the expert is provided an introductory screen with instructions on how to interact with the assessment tool. After the expert clicks on a "proceed" tab, the next screen pops up; this is illustrated in Figure 10. Get ready. Behind that screen, the next screen has been loaded. When ready, the expert clicks on the "next question" tab, and this makes the current screen disappear, revealing the underlying screen that offers the first assessment and is illustrated in Figure 11. Assessment one.

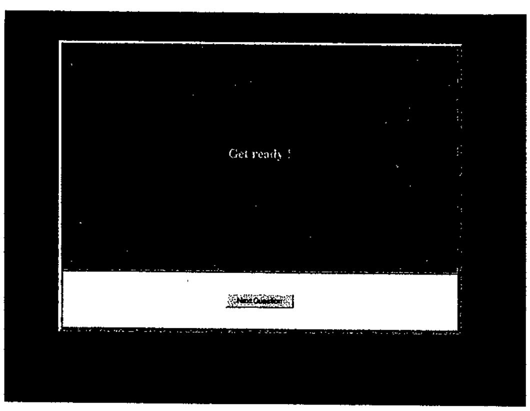

Figure 10. Get ready

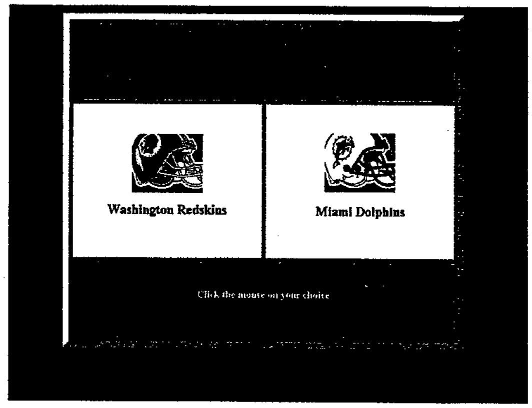

Figure 11. Assessment one

This starts a timer within the expert's browser. Then the expert clicks anywhere in the region of the preferred choice. This stops the timer, captures both the decision time and choice, and sends the results back to the server for storage and processing.

Immediately the screen illustrated in Figure 10. Get ready, returns to the screen, and the process is repeated with all choices. Since there are only three teams in this trivial example, only two more figures are shown. Figures 12 and 13, labeled Assessment two and Assessment three, respectively.

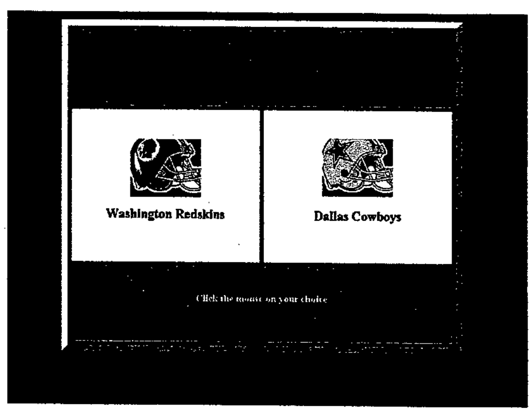

Figure 12. Assessment two

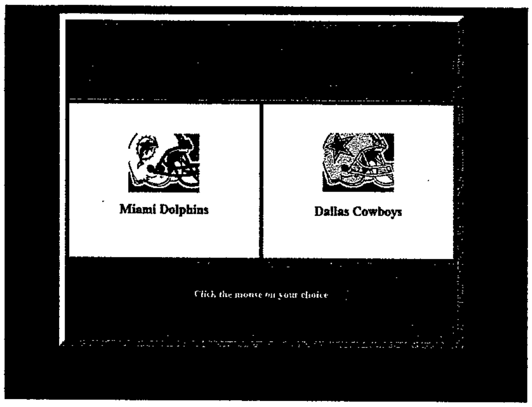

Figure 13. Assessment three

After the final assessment, a message notifies the expert that the analysis is complete and offers to provide the results. This is an option, and in many cases, the results are not provided at the time of completing the assessment.

CONCLUSION

In conclusion, an overview of the process is illustrated in Figure 14. Process overview.

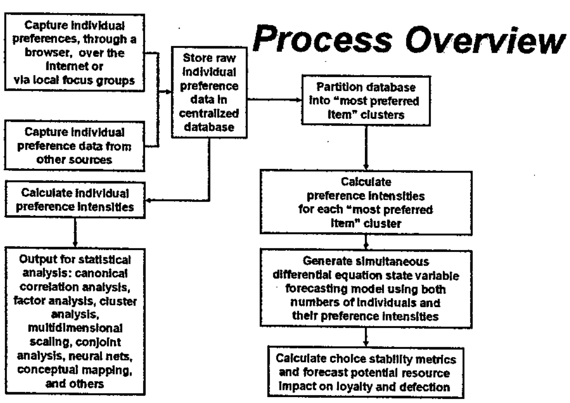

Figure 14. Process overview.

This flow diagram shows how the different parts of the technique may be used in conjunction with other methods. In particular, note that the right side of the diagram refers to the stability metric, and that technique can be used with the response latency and AHP technique or with any other technique that captures preferences, opinions or degrees of confidence.

Current interests involve expanding the decision environment to provide a virtual reality interface to enable analysts to get the feel of flying through the data when performing exploratory analysis as illustrated in Figure 15. The virtual reality interface.

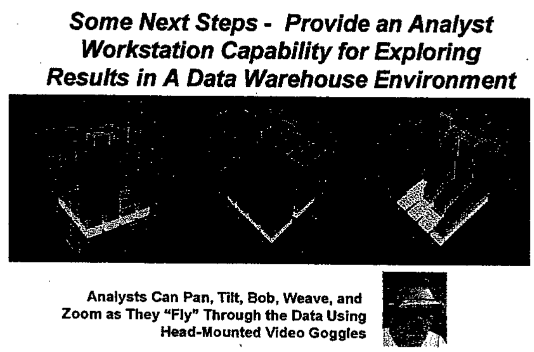

Figure 15. The virtual reality interface.

BIBLIOGRAPHY

Banagi, MR, and Greenwald, A G [1994]. Implicit sterotyping and prejudice. In M. Zanna & J. Olson [Eds.], The Psychology of Prejudice: The Ontario Symposium 7, 55-76. Hillsdale, NJ: Erlbaum. [1995].

Cartwright, D [1941] Decision-time in relation to differentiation of the phenomenal field.
Psychological Review, 48, 425-422.

Clemen, R T [1996] Making Hard Decisions; Duxbury Press, 265-298

Crosby, F, Bromley, S, & Saxe, L [1980]. Recent unobtrusive studies of black and white discrimination and prejudice: a literature review. Psychological Bulletin, 87, 546-563.

Dashiell, JF [1937]. Affective value-distances as a determinant of esthetic judgment-times. American Journal of Psychology, 50, 57-67

Dovidio, JF, and Fazio, R H [1992]. New technologies for the direct and indirect assessment of attitudes. In J. Tanur [Ed.], Questions about Questions: Inquiries into Cognitive Bases of Surveys 204-237. New York: Russell Sage Foundation.

Drucker, P (2001), "The Next Society," *The Economist*, 3 November, 3.

Festinger, L [1943] Studies in decision. Journal of Experimental Psychology, Vol. 32, No. 4, 411- 423.

Forman, E [1999] Personal communications with Professor Ernest Forman, The George Washington University, School of Business and Public Management, Washington, DC.

Gaertner, SL, and Dovidio, J. F. [1986]. The aversive form of racism. In J. F. Dovidio & S. L. Gaertner [Eds.], Prejudice, Discrimination, and Racism 61-89. Orlando, FL: Academic Press.

Glimcher, PW (2001), "Making choices: the neurophysiology of visual-saccadic decision making,". Trends Neurosci, 24:654-659.

Gold JI, Shadlen MN (2001), "Neural computations that underlie decisions about sensory stimuli," Trends Cogn Sci, 5:10-16.

Harmon, P and King, D [1985] Expert Systems - Artificial Intelligence in Business; John Wiley & Sons, Inc. 41-43.

Kahneman, D, Slovic, P., and Tversky, A. [1999] Judgement under uncertainty: Heuristics and biases; Cambridge University Press

LaBarbera, PA, & MacLachlan, JM [1979]. Response latency in telephone interviews. Journal of Advertising Research, 19, 49-56.

Lam, A [1998] Tacit knowledge, organisational learning and innovation: a societal perspective. DRUID Working Paper No. 98-22, Danish Research Unit for Industrial Dynamics

Marshall, KT and Oliver, M R [1995] Decision Making and Forecasting, McGraw-Hill, Inc. 240- 241.

Medsker, L [1998]. Personal communications with Professor Larry Medsker, American University, Department of Computer Science and Information Systems.

Nisbett, RE, and Wilson, TD [1977]. Telling more than we know: verbal reports on mental processes. Psychological Review, 84, 231-259.

Platt, ML (2002), Neural correlates of decisions. Current Opinion in Neurobio, 12:141-148.

Polanyi, M [1962] Personal knowledge: towards a post-critical philosophy. New York: Harper Torchbooks. [1966] The tacit dimension. New York: Anchor Day Books.

Romo R, Salinas E (2001), "Touch and go: decision-making mechanisms in somatosensation,"Annual Rev Neurosci, 24:107-137.

Saaty, TL [1977]. A scaling method for priorities in hierarchical structure. Journal of Mathematical Psychology 15 79-84

_________ [1994] Fundamentals Of Decision Making And Priority Theory with the Analytic Hierarchy Process, Volume IV, 1994, RWS Publications, Pittsburgh, PA., 8-127.

_________[1995] Decision Making for Leaders - The Analytic Hierarchy Process for Decisions in a Complex World. Third Edition. RWS Publications, Pittsburgh, PA., 70-92

[1999] Personal communications with Professor Thomas L. Saaty, the University of Pittsburgh.

Schall JD (2001), "Neural basis of deciding, choosing and acting," Nat Rev Neurosci, 2:33-42.

Schall JD, (2003), "Neural correlates of decision processes: neural and mental chronometry," Current Opinion in Neurobio, 13:182–186

Sigal, H, and Page, R [1971]. Current stereotypes: a little fading, a little faking. Journal of Personality and Social Psychology, 19, 247-255.

Tyebjee, T T, [1979]. Response latency: a new measure for scaling brand preference. Journal of Marketing Research, 16, 1, 96-101.

Zaltman, G, [2003]. How Customers Think – Essential Insights Into the Mind of the Market. Harvard Business School Press. 40-43.